**Методы машинного обучения.**

**ЛР2. - Модель полиномиальной регрессии - Регуляризация.**

**Задание**

Для модели полиномиальной регрессии, полученной в лабораторной работе 1 пункт 1 (оптимальный вариант), а также для полиномов меньшей и большей степеней (±3) вывести значения коэффициентов полинома (всего 3 полинома).

К выбранным моделям (полиномам соответствующих степеней) применить метод регуляризации с использованием гребневой регрессии (ридж-регрессии) и Лассо-регрессии. Вывести значения коэффициентов полинома. Повторить для различных значений параметра *λ.*

Рассчитать функционал эмпирического риска (функционал качества) для всех полученных моделей на обучающей и контрольной выборках (вывести графики).

**Теория**

**Регуляризация.**

**Регуляризация** (англ. *regularization*) в статистике и машинном обучении — метод добавления некоторых дополнительных ограничений к условию с целью предотвратить переобучение. Чаще всего эта информация имеет вид штрафа за сложность модели.

Если выбрана излишне сложная модель при недостаточном объеме данных, то в итоге может быть получена модель, которая хорошо описывает обучающую выборку, но не обобщается на тестовую. Переобучение в большинстве случаев проявляется в том, что итоговые модели имеют слишком большие значения параметров. Одним из способов борьбы с негативным эффектом излишнего подстраивания под данные — использование регуляризации, т.е. добавление некоторого штрафа за большие значения коэффициентов у линейной модели. Тем самым запрещаются слишком "резкие" изгибы, и предотвращается переобучение.

Наиболее часто используемые виды регуляризации — и , а также их линейная комбинация — эластичная сеть.

В представленных ниже формулах для эмпирического риска приняты следующие обозначения: - функция потерь, - вектор параметров из модели алгоритма, – неотрицательный гиперпараметр (коэффициент регуляризации).

Если в качестве функционал качества используется сумма квадратов остатков (Residual Sum of Squares- RSS), тогда изначально:

**-регуляризация** (ridge regularization) или регуляризация Тихонова (Tikhonov regularization):

Минимизация регуляризованного cоответствующим образом эмпирического риска приводит к выбору такого вектора параметров , которое не слишком сильно отклоняется от нуля. В линейных классификаторах это позволяет избежать проблем мультиколлинеарности и переобучения.

**-регуляризация** (*lasso regularization*) или регуляризация через манхэттенское расстояние:

Данный вид регуляризации также позволяет ограничить значения вектора. Однако, к тому же он обладает интересным и полезным на практике свойством — обнуляет значения некоторых параметров, что в случае с линейными моделями приводит к отбору признаков.

#### Указанные регрессии реализованы в библиотеке Scikit-learn

***# Ridge***

from sklearn.linear\_model import Ridge

ridge\_regression = Ridge(alpha=0.1) # alpha — величина регуляризации

*# обучение*

ridge\_regression.fit(train\_X, train\_y)

*# предсказание результата*

*print*(ridge\_regression.predict(test\_X))

*# вывод точности предсказания*

*print*(ridge\_regression.score(test\_X, test\_y))

***# Lasso***

from sklearn.linear\_model import Lasso

lasso\_regression = Lasso(alpha=0.1) # alpha — величина регуляризации

*# обучение*

lasso\_regression.fit(train\_X, train\_y)

*# предсказание результата*

*print*(lasso\_regression.predict(test\_X))

*# вывод точности предсказания*

*print*(lasso\_regression.score(test\_X, test\_y))